

CA 1 Group 3

1. (a) Find a closed-form solution for this problem;

The closed form solution is correct but they have not provided the derivation of the closed form solution.

2. (b) Consider “Individual household electric power consumption” dataset ($N = 2075259$, $d = 9$) and find the optimal linear regressor from the closed-form expression;

- The data has not been checked for nan or any other string value.
- The group has Input
`"Global_reactive_power", "Voltage", "Global_intensity"`
- Output:
`"Global_active_power"`

- The group has used the function (`linear_model.Ridge`) for calculations but have also given in comments how they could have implemented the closed form solution.
- The value of lambda is taken arbitrarily and could have been $(1/N)$.
- The solution seems correct as they have shown with mse that their predicted output is close to the actual one.

3. (c) Repeat 2) for “Greenhouse gas observing network” dataset ($N = 2921$, $d = 5232$) and observe the scalability issue of the closed-form expression;

- The group has solved the problem but has some minor issues:
- First they have taken the data as two groups and have not merged the two data sets, they have two sets of values. The scalability issues could have been much demonstrated if we take large data at one and not in parts.
- The time taken is greater as expected but mse is very large.

4. Iterative method is the correct answer for part d.